IMAGE PROCESSING AND CLASSIFICATION TECHNIQUES IN HEALTHCARE APPLICATIONS

Nalamwar Sonali [1] | Kharabi Kirti [1] | Mahajan Vrushali [1] | Shinde Anushka [1] | [1] Sisode Prapti

[1] Department of Computer Engineering, All India Shree Shivaji Memorial Society's College Of Engineering, Savitribai Phule Pune University, Maharashtra, India.

ABSTRACT

Lung Cancer is leading cause of death in world. Different type of diseases leads to death but it is observed that most of the times death is due to cancer. If cancer is detected in early stage it is helpful in curing cancer completely. Lung cancer is generally misdiagnosed. Image processing and data mining found numerous applications in scientific and healthcare domain. To find out affected part by comparing CT scan image of both normal and affected person, Image processing technique such as smoothing, filtering, enhancement, segmentation, feature extraction are applied. Preprocessing techniques such as smoothing, enhancement and segmentation are applied on the image. Then features such as area, perimeter, eccentricity, curve, edges are extracted from pre-processed image using SIFT algorithm and then decision tree and SVM classifiers are used for classification. Based on classification, stage of cancer can be identified. SVM and decision tree classifiers are used to increase accuracy of the system.

KEYWORDS: Computer Tomography, Decision tree, Image processing, Scale Invariant Feature Transform, Support Vector Machine.

I. INTRODUCTION

Cancer is a general term used to refer to a condition where the body cells begin to grow and reproduce in an uncontrollable way. These cells can then invade and destroy healthy tissue, including organs. Cancer sometimes begins in one part of the body before spreading to other parts. Cancer is a common condition and a serious health problem. More than one in three people will develop some form of cancer during their lifetime. Excluding non melanoma skin cancer, there are around 7,000 new cases diagnosed each year. The figure below shows death rate of lung cancer per 100000 population.
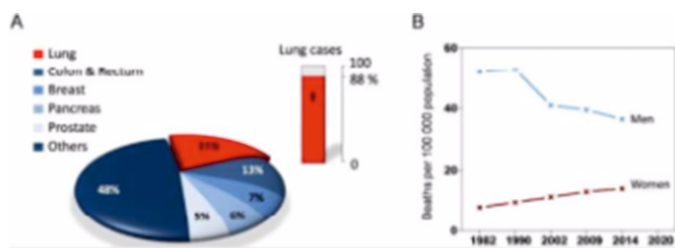

LITERATURE SURVEY

- Lung cancer nodule detection at early stage using SVM Classifier has been proposed . A comparison of classification accuracy for ANN,KNN and SVM Classifiers was made on Lung CT scan images of stage I and stage II[1].

- Neural Networks and SVM for detection of lung cancer in X-ray chest films was used. High number of false positives extracted and a set of 160 features was calculated and feature extraction technique was applied to select the best feature[2].

- Comparison is made between PET and CT to know which gives the best result through applying some image processing techniques. In proposed system, the system design is made for detecting the lung cancer in early stage using SVM Classifier[3].

- Semi supervised classifier are used to classify the nodule and the performance are terms of sensitivity, accuracy, specificity, precision and recall.FCM

- Clustering algorithm was used for segmentation of image and anisotropic diffusion for removing noise[4].

- Feature extraction and neural network classifier is used to check the state of patient in its early stage and to predict survival rate and year of abnormal lung by extracted features from CT image[5].

- The proposed technique gives very promising results comparing with other used techniques. Relying on general features, a normality comparison is made. A hybrid technique based on feature extraction and Principal Component Analysis(PCA) is presented for lung cancer detection in CT scan images[6].

- Comparison of the classification techniques which includes CART, Random Forest, LMT, and the Naive Bayesian over different cancer survival data set is done and it showed that Random forest method using training dataset outperforms the other methods. Relative absolute error of LMT is high for cancer survival dataset[7].

II. PROPOSED ARCHITECTURE

In a proposed system as shown in figure below we need to take a CT scan image of lung as an input to the system. The CT scan image contains noise and has to be processed to get the feature of lung s that classification can be done using these features. The first step of our system is image pre-processing. Image pre-processing includes de-noising i.e. removing the unwanted noise from the image. De-noising is nothing but smoothing.
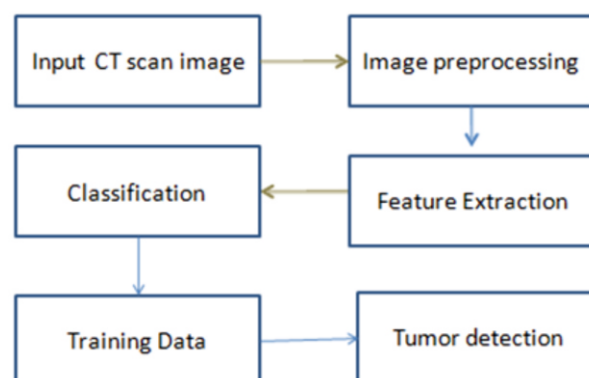
Proposed System Architecture

2.1. Gaussian Filter

Gaussian filter is one of the filtering technique used for de-noising the image. In image processing two dimensional Gaussian function is used.

2.2. Feature Extraction

Scale Invariant Feature Transfom SIFT Algorithm :

1. **Scale-space extrema detection:** First points of interest are detected, which are termed keypoints in the SIFT. The image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken. This is done by comparing each pixel in the DoG images to its eight neighbours at the same scale and nine corresponding neighbouring pixels in each of the neighbouring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate keypoint.

2. **Keypoint localization:** Scale-space extrema detection produces too many keypoint candidates, some of which are unstable. The next step in the algorithm is to perform a detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

3. **Interpolation of nearby data for accurate position:** For each candidate keypoint, interpolation of nearby data is used to accurately determine its position. This approach calculates the interpolated location of the extremum, which improves matching. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function is used.

4. **Eliminating edge responses:** The DoG function will have strong responses along edges, even if the candidate keypoint is not robust to small amounts of noise. Therefore, in order to increase stability, we need to eliminate the keypoints that have poorly determined locations but have high edge responses.

5. **Orientation assignment:** In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation as the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.For an image sample, the gradient magnitude and orientation are computed using pixel difference.

    Keypoint descriptor: Previous steps found keypoint locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. Now we want to compute a descriptor vector for each keypoint such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the keypoint's scale.

6. **Keypoint Matching:** Keypoints between two images are matched by identifying their nearest neighbours. But in some cases, the second closest match may be very near to first. In such case, ratio of closest distance to second closest distance is taken. If it iss greater than 0.8 , they are rejected. It eliminates around 90per cent of false matches while discards only 5 per cent correct matches.

2.3. Classification Techniques

a) Decision tree

C4.5 constructs a classifier in the form of decision tree. For this purpose c4.5 is given a data set which is already classified. Hence C4.5 is supervised learning algorithm. C4.5 classifies is a tool in data mining that takes a bunch of data representing thing which are to be classified

and attempts to predict which class the new data belongs to.DT is like flowchart to classify new data. Using patients attribute information, one particular path in the flowchart could be tumour in lungs, size of tumour greater than 5cm. DT is supervised learning algorithm, since the training dataset already labelled with classes. C4.5 doesnt learn on its own that a patient get cancer or not. Firstly it generate a decision tree on training data set and then it uses this DT for classification.

b) Support Vector Machine(SVM)

Support vector machine is a supervised machine learning algorithm. It does the classification by constructing an n-dimensional hyperplanes which actually segregates data into two partitions. It is a binary classifier in which data parts are classified into classes by using labels i.e. members of the same class have same label. In SVM machine learning is done by set of input values with associated output values. It uses maximum margin

value to separate classes. Use of max margin value reduces the chances of making error. Support vectors are input vectors that touch the boundary of the margin. Support vector are the elements in training data set that may change the position of dividing hyperplane if removed.

SVM also allows non linear mapping if data is not linearly separable, for this it uses non-linear kernel by constructing of new feature space.
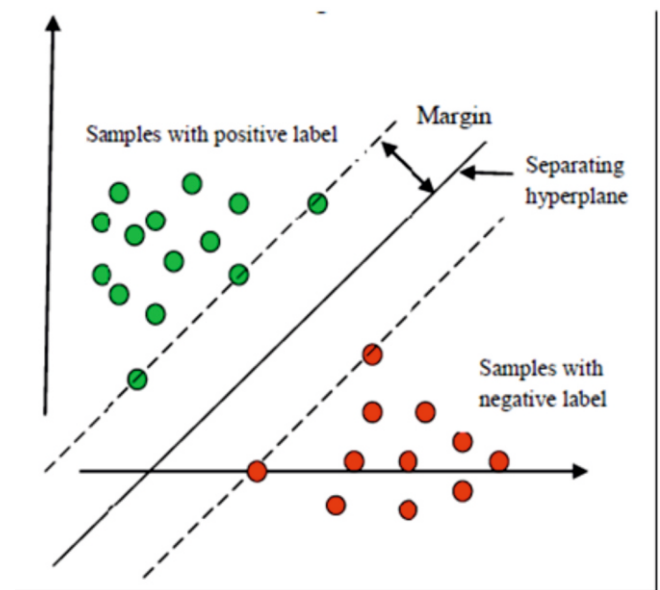
SVM Classifier

III. CONCLUSION

3.1 Conclusions

This proposed system identifies and detects lung cancer based on feature extraction and classification on CT scan images. In this system we will achieve the purpose of developing an automated system which will detect lung cancer. It is useful to detect cancer in early stages which will help in increasing the survival rate.

3.2. Future Works

In future, same work can be done on MRI images and X-ray images. All these images can be compared so as to justify which types of images gives better result for lung cancer detection using different classification techniques.

REFRENCES:

1. P.Nithya, B. Umamaheshwari, R. Deepa,Detection of Lung Cancer Using Data Mining Classification Techniques, International Journal of Advanced Reaserch in Computer Science and software Engineering, Vol 5 pp 1060-1062 July 2015 .

2. Fatehgath Sahib ,A study of Detection of Lung Cancer Using Data Mining Classification Techniques, International Journal of Advanced Reaserch in Computer Science and software Engineering, Vol 3, pp 131-134 March 2013.

3. P.Nivetha, Mr.R.Manickavasagama, Lung Cancer Detection at Early Stage Using PET/CT Imaging Technique, International Journal of Innovative Research in Computer and Communication Engineering, vol 2, March 2014.

4. S.Ramya Preethi, Prof.R.Vijayalakshmi, P.Deepa, , International Journal of Innovative Research in Computer and Communication Engineering, vol 4, May 2015.

5. Ada, Rajneet Kaur, Using Some Data Mining Techniques to Predict the Survival Year of Lung Cancer Patient, International Journal of Computer Science and Mobole Computing,vol 2,pg.1-6,April 2013.

6. Ada, Rajneet Kaur,Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images, International Journal of Innovative Research in Computer and Communication Engineering, vol 3, March 2013,pp 187-190

7. Ankita Kumar, Mr. Mohamed Amanulla,A study of Cancer Perpetuation using the Classification Algorithms, International Journal of Computer Science and Mobile Computing,vol 4,April 2015,pg.395-400.

8. Mickias Assefa, Ibrahima Faye, Aamir Saeed Malik and Muhammad Shoaib, Lung Nodule Detection Using Multi- Resolution Analysis, Proceedings of International Conference on Complex Medical Engineering, pp. 457-461,2013.

9. Nguyen,H.T.,et al ,Watersnakes : Energy-Driven Watershed Segmentation, IEEE Transactions on pattern Analysis and Machine Intelligence ,Volume 25, Number 3,pp.330-342, March 2013.

10. Disha Sharma, Gagandeep Jindal, Identifying Lung Cancer Using Image Processing Techniques, Internationalconference on computational techniques and artificial intelligence (ICCTAI), pp. 115-120, 2011.

11. Ada, Rajneet Kaur, Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, pp. 187-190, Mar.2013.

12. Stelmo Magalhaes BarrosNetto, Aristo fanes Correa Silva, Rodolfo Acatauassu Nunes and MarceloGattass, Automatic segmentation of lung nodules with growing neural gas and support vector machine, Computers in Biology and Medicine, pp. 1110-1121,Sep 2012.

13. Henry Krewer, Benjamin Geiger Lawrence O. Hall, Dmitry B. Goldgof, Yuhua Gu, Melvyn Tockman and Robert J. Gillies, Effect of Texture Features in Computer Aided Diagnosis of Pulmonary Nodules in low Dose Computed Tomography, IEEE International Conferenceon Systems, Man, and Cybernetics, pp. 3887-3891,2013.

14. Anita chaudhary, Sonit Sukhraj Singh, Lung cancer detection on CT images using Image processing, International Conference on Computing Sciences, pp.142- 146,2012.

15. Archana S, K.P. Kaliyamurthie and V.Khanaa, CAD System for Lung Cancer Detection, International Journal Of Engineering And Computer Science, Vol. 2 , pp. 921-925, Apr.2013.

16. J. David Schaffer, Jin Woo Park, Erin Barnes, Qiyi Lu, Xingye Qiao, Youping Deng Yan Li , Walker H. Land, GRNN ensemble classifier for Lung Cancer prognosis using only demographic and TNM features, SciVerse Science Direct, pp. 450-455,2012.

17. Mr.Vijay A.Gajdhane , Prof. Deshpande L.M. , Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques, IOSR Journal of Computer Engineering (IOSR-JCE).

18. G. Vijaya, A. Suhasini, R. Priya , Automatic Detection Of Lung Cancer In CT Images, International Journal of Research in Engineering and Technology

19. Ada, Rajneet Kaur, Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier, International Journal of Application or Innovation in Engineering and management.